8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.2, June 6-8, 2007, Cancún, Mexico

# IDENTIFYING CHEMICAL REACTION NETWORK MODELS

**S. C. Burnham, M. J. Willis and A. R. Wright**

School of Chemical Engineering and Advanced Materials,
Newcastle University, Newcastle-upon-Tyne, NE1 7RU, UK.

Abstract: This paper demonstrates the identification of underlying nonlinear ordinary differential equation (ODE) models that represent chemical reaction networks. The proposed method uses species' concentration and rate of heat evolution data which can be obtained from a reaction calorimeter. The identification procedure is demonstrated for a simulated Van de Vusse reaction system. It is shown that accurate estimates of the network rate constants and individual heats of reactions may also be generated. 

Keywords: Chemical industry, Reactor modelling, Differential equations, Structural optimisation, System identification.

## 1. INTRODUCTION

One of the key issues for fine chemical and pharmaceutical companies is to reach the market with many new products as quickly as possible. Within these industries multi-purpose plants which are suitable for a variety of customer specifications are often used. This leads to fast changing discontinuous processes incorporating batch or semi-batch reactors. A major problem facing these companies is the scale up of a process from laboratory to full-scale production in the shortest possible time, preferably avoiding pilot-plant testing.

Modelling and simulation studies are often used to assist scale-up. For conventional equation-based modelling software it is necessary to formulate mathematical equations in order to describe the process dynamics. In general, the behaviour of process design variables such as flow-rates, reactor volumes and inlet concentrations of species are well understood. However, obtaining knowledge of the chemistry, in particular the chemical reaction network remains a limiting step.

Traditional methods for determining reaction networks involve postulating a number of different reaction networks which are then fit to experimental data. The reaction network whose model provides the highest prediction accuracy with respect to the experimental data is taken to be the correct structure. This is a time consuming procedure requiring both chemistry and modelling expertise.

More recently a semi-automated procedure has been proposed (Burnham *et al.* 2006, 2007). This method involves specifying a global ODE model structure capable of representing the entire set of possible chemical reactions. Mathematical and statistical tests are then used to reduce the ODE model structure to a subset of reactions, which can be combined to give the reaction network. However, the disadvantage of this approach was the need for a model rationalisation procedure relying on exploiting the basic rules of reaction chemistry. This model rationalisation procedure was not automated and relied on the judgment of the user.

In this paper the method is modified in order to develop a less subjective and more automated procedure. It is also extended to incorporate the rate of heat evolution ($Q_r$) data. The benefit of this approach is demonstrated using a simulated Van de Vusse reaction network.

## 2. CHEMICAL REACTION NETWORKS AND RATE EXPRESSIONS

Chemical reaction networks are composed of a number of elementary reactions. For example, consider the reaction network (1).

$$\begin{array}{l} x_1 + x_1 \xrightarrow{k_1} x_2 \xrightarrow{k_2} x_3 \\ x_1 + x_2 \xrightarrow{k_3} x_4 + x_5 \end{array} \quad (1)$$

The network involves five chemical species, $x_1$, $x_2$, …, $x_5$. There are three elementary chemical reactions a) an addition reaction of $x_1$ with itself to form $x_2$, b) $x_2$ rearranges to form $x_3$, c) an addition reaction of $x_1$ and $x_2$ to form the two products $x_4$ and $x_5$. $k_1$, $k_2$ and $k_3$ are the rate constants (for an isothermal system).The reaction rates $r_i$, describe the rate of consumption or formation of each species in the $i^{th}$ reaction. Under the assumption of mass action kinetics, for network (1) the reaction rates are shown by (2).

$$r_1 = 2k_1[x_1]^2 \qquad r_2 = k_2[x_2] \qquad r_3 = k_3[x_1][x_2] \quad (2)$$

Where $[x_1]$, $[x_2]$ are the respective species concentrations. For the purposes of this paper the set of concentration terms $\{[x_1]^2, [x_2], [x_1][x_2]\}$ that appear within the rate expressions (2) will be referred to as complexes.

## 3. IDENTIFICATION OF ODE MODELS THAT REPRESENT CHEMICAL REACTION NETWORKS

In a constant volume batch reactor the rate of change of each species at any time may be determined by a set of ODEs. These ODEs must be physically consistent with each other in the sense of conservation of mass.

For network (1) the set of ODEs are shown by (3).

$$\begin{array}{lll} \frac{\mathrm{d}[x_1]}{\mathrm{dt}} = -r_1 - r_3 & \frac{\mathrm{d}[x_2]}{\mathrm{dt}} = \frac{r_1}{2} - r_2 - r_3 & \\ \frac{\mathrm{d}[x_3]}{\mathrm{dt}} = r_2 & \frac{\mathrm{d}[x_4]}{\mathrm{dt}} = r_3 & \frac{\mathrm{d}[x_5]}{\mathrm{dt}} = r_3 \end{array} \quad (3)$$

For an unknown reaction network, not only must the correct rate terms be selected for each ODE but the rate constants and complexes for each rate term must also be determined. Hence the task of identifying a useful network is one of determining the structure as well as the parameters of a set of non-linear ODEs. This is a non-trivial system identification task.

To make this identification task tractable it may be partitioned into simpler subtasks. Firstly, the number of reactions, i.e. the number of rate expressions, is estimated. Secondly, best subsets regression against the $Q_r$-time profile data is performed to identify the concentration complexes associated with each rate expression. Thirdly, an additional regression stage is used to determine the structure of the individual species' ODEs. Finally, the kinetic rate constants and the individual heats of reactions are determined using standard kinetic fitting software.

## 4. DETERMINING THE NUMBER OF INDEPENDENT REACTIONS

Bernard and Bastin (2005) and Brendel *et al*. (2006) describe an approach to estimate the number of reactions in a biotechnological process. This method is applicable to data from a constant volume, isothermal batch reactor. If it is assumed that a) all species are measured b) the number of reactions is less than the number of species, then the number of independent reactions may be determined through assessment of the linear dependence of the ODEs. Let $\boldsymbol{X}$ be defined as a matrix with column vectors that represent the individual species ODEs. Then the rank of $\boldsymbol{X}$ is equal to the number of linearly independent columns and this rank is the number of independent reactions, $l$. For the ODEs described by (3) $\boldsymbol{X}$ is given by (4).

$$\boldsymbol{X} = \left[ \overrightarrow{-(r_1 + r_3)}, \quad \overrightarrow{(\frac{r_1}{2} - r_2 - r_3)}, \quad \vec{r}_2, \quad \vec{r}_3, \quad \vec{r}_3 \right] \quad (4)$$

These rate terms are not known in advance, however, the individual species' ODEs may be approximated as the rate of change of concentration of the individual species due to reaction. Hence, it is necessary to estimate a) the rate of change of concentration of the each species b) the rank of matrix $\boldsymbol{X}$.

### *4.1 Estimating the rate of change of concentration.*

The rate of change of concentration of each species may be approximated from the slopes, $S[x_i]_t$, of the measured concentration data. These may be calculated for all measured time points $t = 1,\ldots, N$. For network (1), this allows $\boldsymbol{X}$ to be approximated to the matrix $\hat{\boldsymbol{X}}$ defined by (5).

$$\hat{\boldsymbol{X}} = \left[ \overrightarrow{S[x_1]} \quad \overrightarrow{S[x_2]} \quad \cdots, \quad \overrightarrow{S[x_5]} \right] \quad (5)$$

Obtaining a good approximation of the first derivatives for the concentration of each chemical species is important to the success of the method. Several options are available for determining the slope of the measured concentration data. For example, Almeida and Voit (2003) and Voit and Almeida (2004) have demonstrated that with artificial neural networks it is possible to obtain sufficiently accurate approximations to the derivatives, for the purposes of determining useful biochemical pathway information. Burnham *et al.* (2007) describe the fitting of rational polynomials to the concentration profiles in order to obtain estimated derivatives.

*4.2 Determining the rank of a matrix.*

In the presence of noise, the matrix $\hat{\boldsymbol{X}}$ will be full rank. However, it is possible to decompose $\hat{\boldsymbol{X}}$ using the singular value decomposition (SVD) given by (6).

$$\hat{\boldsymbol{X}} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} \tag{6}$$

Where **U** is a matrix of the eigenvectors of $\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^{T}$ and **V** is a matrix of the eigenvectors of the data correlation matrix, $\hat{\boldsymbol{X}}^{T}\hat{\boldsymbol{X}}$. The elements of the diagonal matrix Σ are the positive square roots of the eigenvalues, $\lambda_i$, of $\hat{\boldsymbol{X}}^{T}\hat{\boldsymbol{X}}$ and are called the singular values. The eigenvalues correspond to the variance associated with the corresponding eigenvector. With SVD the singular values are normally arranged in descending order. The largest singular value will explain the most variation in the data set; the second explains the next level of variation etc. To determine the number of linearly independent columns, it is necessary to determine the number of singular values that must be taken into account. The generally accepted method consists of selecting the set of largest singular values which represent a summed variance larger than a fixed threshold. The number of singular values, $\hat{l}$, contained in this set represents the estimated rank of matrix $\hat{\boldsymbol{X}}$ hence the estimated number of independent reactions.

## 5. DEFINING THE SET OF POSSIBLE COMPLEXES

Once the estimated number of reactions (and hence, number of rate expressions) $\hat{l}$ is known, it is necessary to define the structure of the individual rate expressions. This may be achieved by, firstly, specifying a set of possible complexes that may be used to construct the rate expressions. If it is assumed that a) the reactions are at most the result of bimolecular collisions and the total reaction order is no greater than two b) the complexes are of integer order with respect to each species concentration c) there are no terms consisting of purely zero order elements. Then, given the $n$ species, all possible complexes are given by (7).

$$\begin{aligned} &[x_i] && i = 1,\cdots,n \\ &[x_i][x_j] && i = 1,\cdots,n;\ \ j = i,\cdots,n \end{aligned} \tag{7}$$

Where the total number of possible complexes, $m$, is given by (8).

$$m = 2n + \frac{n(n-1)}{2} \tag{8}$$

Whilst (7) represents the set of all possible complexes, the number of actual complexes that comprise the individual rate expressions is $\hat{l}$ (the estimated number of independent reactions). One way to determine these $\hat{l}$ complexes is to regress against the $Q_r$-time profile data obtained from a reaction calorimeter.

## 6. CALORIMETRY

Regenass (1997) defines two types of calorimetry methods, namely heat accumulation and heat flow. In the heat flow method, feedback control is used to compensate for any temperature changes, maintaining isothermal conditions or following a predefined temperature profile. The power required to achieve isothermal conditions equates to the total heat generated by the reactions.

With the heat flow method, $(Q_r)_t$, is equivalent to the sum of individual heats of reaction multiplied by the rates of reaction. For $l$ reactions $(Q_r)_t$ (kJs$^{-1}$) is given by (9).

$$(Q_r)_t = \sum_{i=1}^{l} (r_i)_t (-\Delta H_i) \tag{9}$$

Where $(-\Delta H_i)$ is the heat of $i^{th}$ chemical reaction (kJkmol$^{-1}$) and $(r_i)_t$ is the rate of the $i^{th}$ chemical reaction at time $t$.

## 7. BEST SUBSETS REGRESSION OF THE $Q_r$-TIME PROFILE DATA

Best subsets regression allows a number of models containing different predictors and different numbers of predictors to be compared. Hence, it is possible to compare models comprised of all possible combinations of $\hat{l}$ terms from the set of $m$ concentration complexes given by (7). Note that this procedure does not produce a regression equation but identifies the best combination of $\hat{l}$ concentration terms that describe the variability in the $Q_r$ expression. The adjusted $R^2$ value and Mallow's *C-p* statistic are used to determine the better model.

The adjusted $R^2$ value is the coefficient of determination describing the relative explanatory power of the overall model (% variability explained by the model), but contains a penalty for models containing many terms. Hence, high adjusted $R^2$ values suggest accurate models.

Mallow's *C-p* statistic is given by (10).

$$C\text{-}p = \left(\frac{SSE_{\hat{l}}}{MSE_m}\right) - \left(N - 2\hat{l}\right) \tag{10}$$

Where $SSE_{\hat{l}}$ is the sum of squared error for the model with $\hat{l}$ particular parameters, and $MSE_m$ is the mean squared error for the model with all $m$ predictors and $N$ is the number of observations. A smaller value of Mallow's *C-p* statistic, of the order of $\hat{l}$, indicates that the model is relatively precise.

Once the best subset of complexes have been identified the structures of the $Q_r$ expression and the $r_i$ expressions are known with respect to the concentration complexes. To elucidate reaction mechanism it is still necessary to determine the combinations of the $r_i$ expressions that comprise the ODEs.

## 8. BEST SUBSETS REGRESSION OF THE ODES

Given that the derivatives of the concentration profiles have been estimated it is possible to identify the complexes (thus, the rate expressions) associated with each ODE. This is achieved by comparing models comprised of all possible combinations of 1 to $\hat{l}$ terms from the subset of concentration complexes determined in section 7. Again this procedure does not produce a regression equation but identifies the best combination of 1 to $\hat{l}$ complexes that describe the variability in each ODE.

## 9. CALCULATION OF THE KINETIC RATE CONSTANTS AND THE INDIVIDUAL HEATS OF REACTION

Once the structure of the ODEs has been determined it is then necessary to accurately estimate the kinetic rate constants. This involves the repeated simulation (integration) of the ODEs whilst adjusting the kinetic parameters using an optimisation algorithm, until the simulation closely matches the experimental data.

The kinetic rate constants are obtained by minimisation of the cost function, given by (11).

$$J = \sum_{i=1}^{n}\sum_{t=1}^{N}([\bar{x}_i]_t - [x_i]_t)^2 \qquad (11)$$

Where the $[x_i]_t$ and $[\bar{x}_i]_t$ are the measured and simulated concentration values respectively. Having obtained the kinetic rate constants, multiple linear regression (MLR) of equation (9) may be completed in order to determine the individual heats of reaction.

## 10. CASE STUDY

The four species Van de Vusse reaction network (Van de Vusse, 1964) is given by (12). In this scheme $x_1$, reacts to produce two products $x_2$ and $x_3$ and $x_3$ further reacts to form $x_4$.

$$x_1 + x_1 \xrightarrow{k_1} x_2 \qquad x_1 \xrightarrow{k_2} x_3 \xrightarrow{k_3} x_4 \qquad (12)$$

The rate expressions are shown by (13).

$$r_1 = 2k_1[x_1]^2 \quad r_2 = k_2[x_1] \quad r_3 = k_3[x_3] \qquad (13)$$

The rate of change of concentration of each species due to reaction is given by (14).

$$\frac{d[x_1]}{dt} = -r_1 - r_2 \quad \frac{d[x_2]}{dt} = \frac{r_1}{2}$$
$$\frac{d[x_3]}{dt} = r_2 - r_3 \quad \frac{d[x_4]}{dt} = r_3 \qquad (14)$$

The $Q_r$ expression is given by (15).

$$Q_r = -r_1(-\Delta H_{r1}) - r_2(-\Delta H_{r2}) - r_3(-\Delta H_{r3}) \qquad (15)$$

For the purpose of generating simulated data, it was assumed that the reaction was performed in a 1litre reactor at 60°C with water as the solvent. The rate constants were assigned as $k_1$ = 1.00e$^{-3}$ min$^{-1}$mol$^{-1}$dm$^3$ and $k_2$ = 6.85e$^{-3}$min$^{-1}$, $k_3$ = 2.48e$^{-3}$ min$^{-1}$. The heats of reaction were assumed to be $\Delta H_{r1}$=-8.0e$^4$, $\Delta H_{r2}$=-1e$^4$ and $\Delta H_{r3}$=-1e$^5$ kJkgmol$^{-1}$. It was assumed that only the initial concentration of the reactant $x_1$ may be altered. Two simulated batch reactions were carried out and the initial concentrations for $x_1$ were arbitrarily chosen to be 0.30 and 0.80moldm$^{-3}$, the initial concentrations of the other species' were specified as 0.00moldm$^{-3}$.

Eleven concentration measurements of each species were generated for both of the simulated batch experiments by integrating the true system ODEs, using the initial reactant concentrations. The simulated batch reactions were run for 20 hours with sampling every 2 hours. Using (13) and (15), $Q_r$ was calculated at the same sampling times.

The concentration measurements and the $Q_r$ values were subjected to additive Gaussian noise with zero mean and a variance equal to 10% of the maximum absolute value of the signal during the experiment. To avoid conditioning problems and give the same weighting to all variables, improving subsequent numeric computations, all data sets were normalised. This is achieved by centring the data at a zero mean and scaling to unit standard deviation. Note that all figures show re-scaled data.

The curve fitting toolbox in MATLAB® v. 7.1 was used to approximate the concentration profiles and in turn obtain accurate estimates of the derivatives. There are a number of potential model structures that may be used, in this case polynomial models were chosen. MLR was completed to fit polynomials to the noisy simulated data using the MATLAB® v. 7.1 curve fitting toolbox. For example, for the second batch (initial concentration of $x_1$ at 0.80moldm$^{-3}$), the simulated noisy concentration data and the fits of this data are given by figure 1.

The derivatives were estimated by differentiating the polynomial expressions, at the specified sampling times giving $\hat{\boldsymbol{X}}$. The estimated first derivatives for the first batch are plotted in figure 2 alongside the actual derivatives for the system. It can be seen that reliable estimates have been achieved for the derivatives. Note that, although the actual derivatives are plotted alongside the estimates for comparison, they would in practice be unknown.

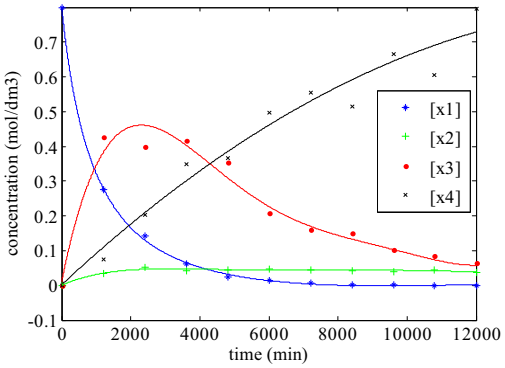


Fig. 1. Concentration-time profiles for a single batch experiment (measured and estimated).

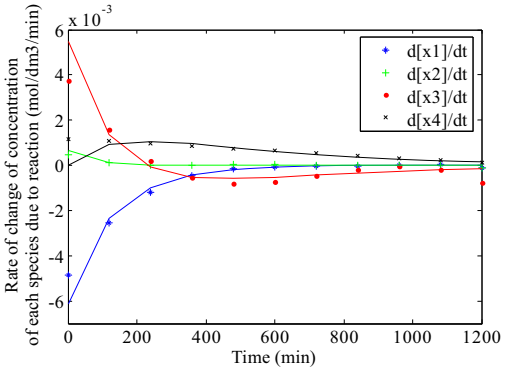


Fig. 2. Actual and estimated rate of change of concentration of species over time for a single batch of measured data.

$\hat{\boldsymbol{X}}$ was decomposed using SVD, to obtain the singular values of $\hat{\boldsymbol{X}}$. The cumulative percentage variance represented by each singular value is shown in figure 3.

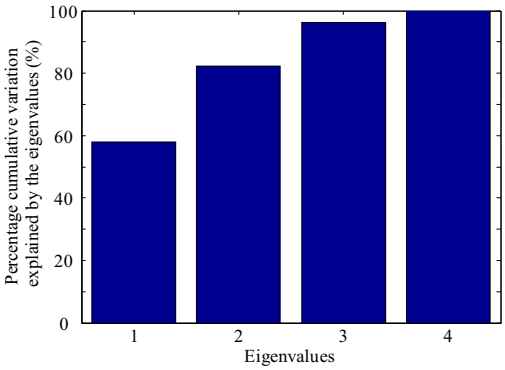


Fig. 3. Cumulative variation explained by the singular values of the matrix $\hat{\boldsymbol{X}}$.

The number of largest singular values which represent a summed variance larger than a fixed threshold of 95% was then determined. This corresponded to the first three singular values which accounted for 96.18% of the variation, hence $\hat{l}$ is three.

As the number of species $n$ is four, the set of fourteen possible complexes are defined by (16).

$$[x_1],\ [x_2],\ [x_3],\ [x_4],\ [x_1]^2,\ [x_1][x_2],\ [x_1][x_3],\ [x_1][x_4],\ [x_2]^2,\ [x_2][x_3],\ [x_2][x_4],\ [x_3]^2,\ [x_3][x_4],\ [x_4]^2 \quad (16)$$

MINITAB®, Release 14.1, was used to perform best subsets regression of the fourteen complexes defined by (16) against the $Q_r$ data. Given that it was estimated that the reaction scheme contained three reactions, the number of predictors for the regression models was specified as three. With an adjusted $R^2$ value of 99.6% and a Mallow's $C$-$p$ statistic of 5.4 the $Q_r$ expression was taken to be proportional to the three complexes, $[x_1]^2$, $[x_1]$, $[x_3]$. Thus the form of the $Q_r$ reaction is given by (17).

$$Q_r = f([x_1]^2, [x_1], [x_3]) \quad (17)$$

From this, the rate expressions are shown by (18).

$$r_1 = 2k_1[x_1]^2 \quad r_2 = k_2[x_1] \quad r_3 = k_3[x_3] \quad (18)$$

In order to determine the reaction network, best subsets regression of models containing one, two and three of the complexes $[x_1]$, $[x_3]$, $[x_1]^2$ was performed against the individual species' ODEs. This suggested that the individual ODEs were proportional to the complexes shown in table 1.

As each ODE expression must be physically consistent with the others in the sense of conservation of mass and adherence to the law of mass action kinetics, it is possible to extract the reaction network by finding matching complex terms in the rate expressions. Through logical application of basic chemical heuristics it can be concluded that $x_1$ is consumed in two ways, through rearrangement to form $x_3$ and reaction with itself to form $x_2$. It can be further concluded that $x_3$ rearranges to produce $x_4$. This gives the reaction network to be (19).

$$x_1 + x_1 \xrightarrow{k_1} x_2 \qquad x_1 \xrightarrow{k_2} x_3 \xrightarrow{k_3} x_4 \quad (19)$$

Given that (19) is the same as (12) the correct reaction network has been constructed.

Table 1 The complexes of each ODE along with the $R^2$ adjusted values and Mallow's $C$-$p$ statistics.

| ODE | Complexes | $R^2$ adj. (%) | Mallow's $C$-$p$ |
|---|---|---|---|
| $\frac{d[x_1]}{dt}$ | $[x_1]^2, [x_3]$ | 99.5 | 2.0 |
| $\frac{d[x_2]}{dt}$ | $[x_1]^2$ | 99.8 | 2.6 |
| $\frac{d[x_3]}{dt}$ | $[x_1], [x_3]$ | 96.4 | 2.5 |

| | | | |
|---|---|---|---|
| $\frac{d[x_4]}{dt}$ | $[x_3]$ | 98.6 | 1.1 |

In this work the BatchCAD v. 8.0 (Copyright © 1995-2004, Aspen Technology, Inc.) was used to fit the kinetic rate constants using a Simplex optimisation algorithm to minimise (11) and an adaptive Runge-Kutta integrator was used to simulate the set of ODEs that represent the reaction network (19). These were fitted to the concentration measurements in order to determine the kinetic rate constants. The rate constants were estimated to be $k_1$ = 9.826$^{-4}$ min$^{-1}$mol$^{-1}$dm$^3$ and $k_2$ = 6.936e$^{-3}$, $k_3$ = 2.494e$^{-3}$ min$^{-1}$, very close to the rate constants specified for simulation.

Having determined the rate constants, regression of the $[x_1]$, $[x_3]$, $[x_1]^2$ terms against $Q_r$ allowed the calculation of the individual heats of reaction. These were calculated as $\Delta H_{r1}$=-9.20e$^4$, $\Delta H_{r2}$=-9.33e$^3$ and $\Delta H_{r3}$=-9.45e$^4$ kJkgmol$^{-1}$, which approximate to those specified for simulation.

The prediction model is plotted against the measured concentration points for a validation batch, figure 4.

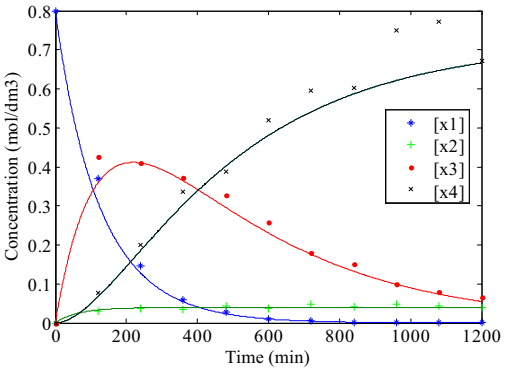


Fig. 4. Prediction of a validation batch profile using the final rationalized model for the Van de Vusse reacion scheme

## 11. CONCLUSIONS

A method for the identification of the underlying nonlinear ODE model representing a reaction network has been presented. A case study using simulated data from an isothermal constant volume batch processes was used to demonstrate the technique. It was shown that it is possible to identify the ODEs that represent the underlying reaction network. Furthermore accurate estimates of the network rate constants and individual heats of reactions were obtained.

Future work will aim to a) test the sensitivity of the threshold value in SVD for determining the number of chemical reactions b) determine the benefit of the use of confidence bounds on the value of $\hat{l}$ c) evaluate the robustness of the mechanism deduction process with respect to unmeasured intermediates and catalysed chemical reactions d) assess of the robustness of the method when $Q_r$ data is unavailable e) apply this technique to real data obtained from a reaction calorimeter.


## ACKNOWLEDGEMENTS

The authors wish to acknowledge the Engineering and Physical Sciences Research Council (EPSRC), UK, for funding this work and K. Novakovic and P. J. English for proof reading the paper.


## REFERENCES


Almeida, J.S. and Voit, E.O. (2003). Neural network based parameter estimation in complex biochemical systems. *Genome Inform*., 14, 114-123.

Bernard, O. and Bastin, G. (2005). On the estimation of the pseudo-stoichiometric matrix for macroscopic mass balance modelling of biotechnological processes. *Mathematical Biosciences,* 193, 51 – 77.

Brendel, M., D. Bonvin, and W. Marquardt (2006) Incremental identification of kinetic models for homogeneous reaction systems. *Chemical Engineering Science*, 61: p. 5404-5420.

Burnham, S.C., Searson, D.P., Willis, M.J. and Wright, A.R. (2006). Towards the automated deduction of chemical reaction mechanism. *Proceedings of the 17th CHISA International Congress of Chemical Engineering, Prague.*

Burnham, S.C., Searson, D.P., Willis, M.J. and Wright, A.R. (2007). Inference of chemical reaction networks from experimental data, *Submitted to Chem. Eng. Sci.*

Regenass, W. (1997). The development of heat flow calorimetry as a tool for process optimization and process safety. *Journal of Thermal Analysis*. 49, 3, 1661-1675.

Voit, E.O. and Almeida, J.S. (2004). Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics*, 20, 11, 1670-1681.

Van de Vusse, J. G. (1964). Plug-flow type reactor versus tank reactor. *Chem. Eng. Sci.,* 19, 994-997.